Analysis Methods for Hadron Colliders II

Beate Heinemann

UC Berkeley and Lawrence Berkeley National Laboratory

TRUMF Summer Institute, Vancouver, July 2009

Outline

- Lecture I:
  - Measuring a cross section
    - focus on acceptance
- Lecture II:
  - Searching for a new particle
    - focus on backgrounds
- Lecture III:
  - Measuring a property of a known particle

Search for New Particles: Experimentally

Number of observed events: counted

Background: Measured from data / calculated from theory

$$\sigma_{NP} = \frac{N_{obs} - N_{BG}}{\int L dt \cdot \varepsilon}$$

Cross section σ

Luminosity: Determined by accelerator, trigger prescale, …

Efficiency: optimized by experimentalist

- Exactly like with measuring the cross section…

But we need to observe first!

- When we don't know if a particle exists our first question is: "Does it exist?"
- => significance of signal
  - I.e. how consistent is the number of observed events with the number of background events?
  - Background expectation: N_{BG}
    - Expect it to fluctuate statistically by $\delta N_{BG} \sim \sqrt{N_{BG}}$
  - Signal expectation: N_{Signal}
  - Statistical Significance: $N_{Signal}/\delta N_{BG} \sim N_{Signal} / \sqrt{N_{BG}}$

in Gaussian limit

    - Often called $S/\sqrt{B}$

	evidence	observation
significance	**3σ**	**5σ**
Probability of stat. fluctuation	**0.3%**	**5.7×10^{-8}**

In real life: systematic uncertainties also contribute to δN_{BG}

Search analyses

- Primary focus is background estimate
  - Determines whether or not an observation can be made
  - Cuts for background reduction studied often using benchmark New Physics scenario
    - Also model-independent analyses attempted sometimes
- Secondary focus is acceptance/efficiency determination: required only
  - when putting an upper limit on a cross section
  - when measuring the cross section of the observed new particle
    - Need to know what it is though
    - Or quote cross section for some effective cuts

Example Analyses

- SUSY:
  - Squarks/gluinos $\rightarrow$ jets + $\not{E}_T$ (+leptons)
- Higgs:
  - Higgs -> WW

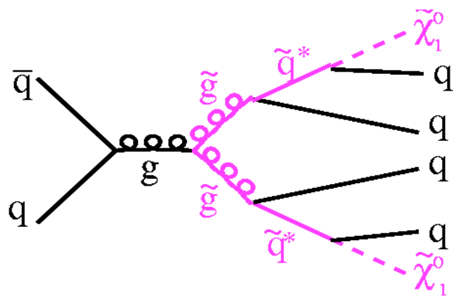
Backgrounds

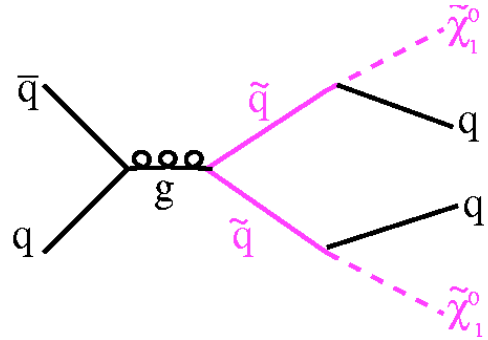
- Ideally you get the backgrounds to be small
  - The smaller they are the less well you need to know them
- Estimates based on
  - Data only
    - E.g. lepton fake rates
  - Monte Carlo only
    - For well known electroweak processes
  - Monte Carlo / Data hybrid
    - For e.g. W/Z+jets or W/Z+b-jets

Squarks/Gluinos $\rightarrow$ Jets + MEt (+ leptons)

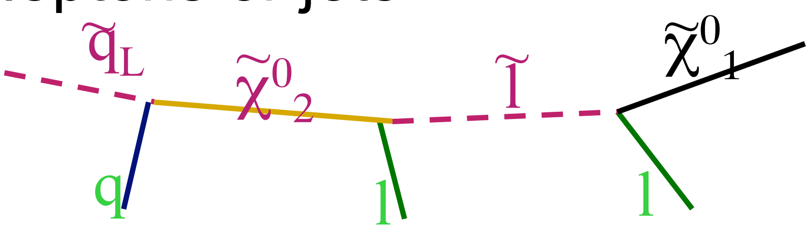
SUSY at the LHC

- Cross section **much** higher than at Tevatron, e.g.
  - for $m(\tilde{g})$=400 GeV: $\sigma_{LHC}(\tilde{g}\tilde{g})/\ \sigma_{Tevatron}(\tilde{g}\tilde{g}) \approx 20{,}000$
  - for $m(\tilde{q})$=400 GeV: $\sigma_{LHC}(\tilde{q}\tilde{q})/\ \sigma_{Tevatron}(\tilde{q}\tilde{q}) \approx 1{,}000$
    - Since there are a lot more gluons at the LHC (lower x)



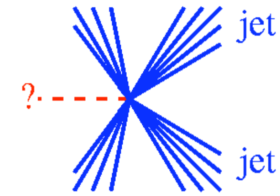


- At higher masses more phase space to decay in cascades
  - Results in additional leptons or jets

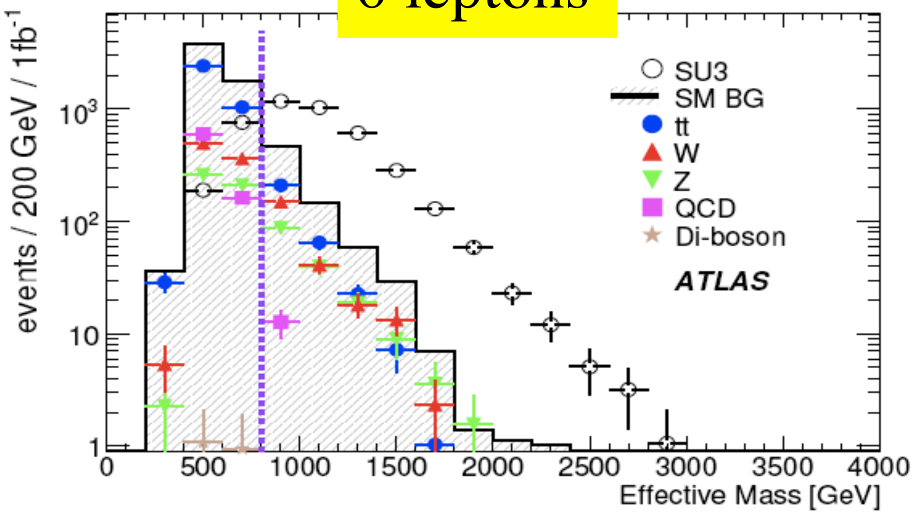


SUSY at the LHC

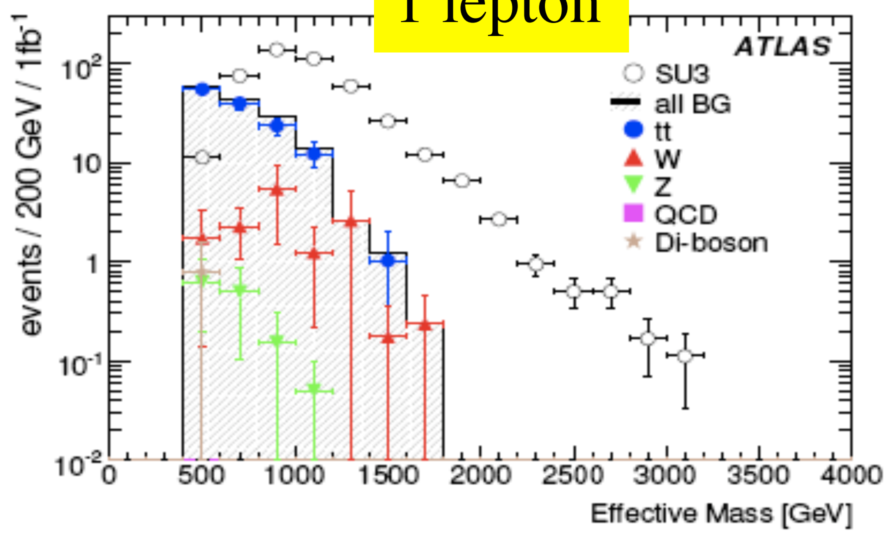
- Example: m($\tilde{q}$)~600 GeV, m($\tilde{g}$)~700 GeV
- Require 4 jets, large missing E_T and 0 or 1 lepton



0 leptons



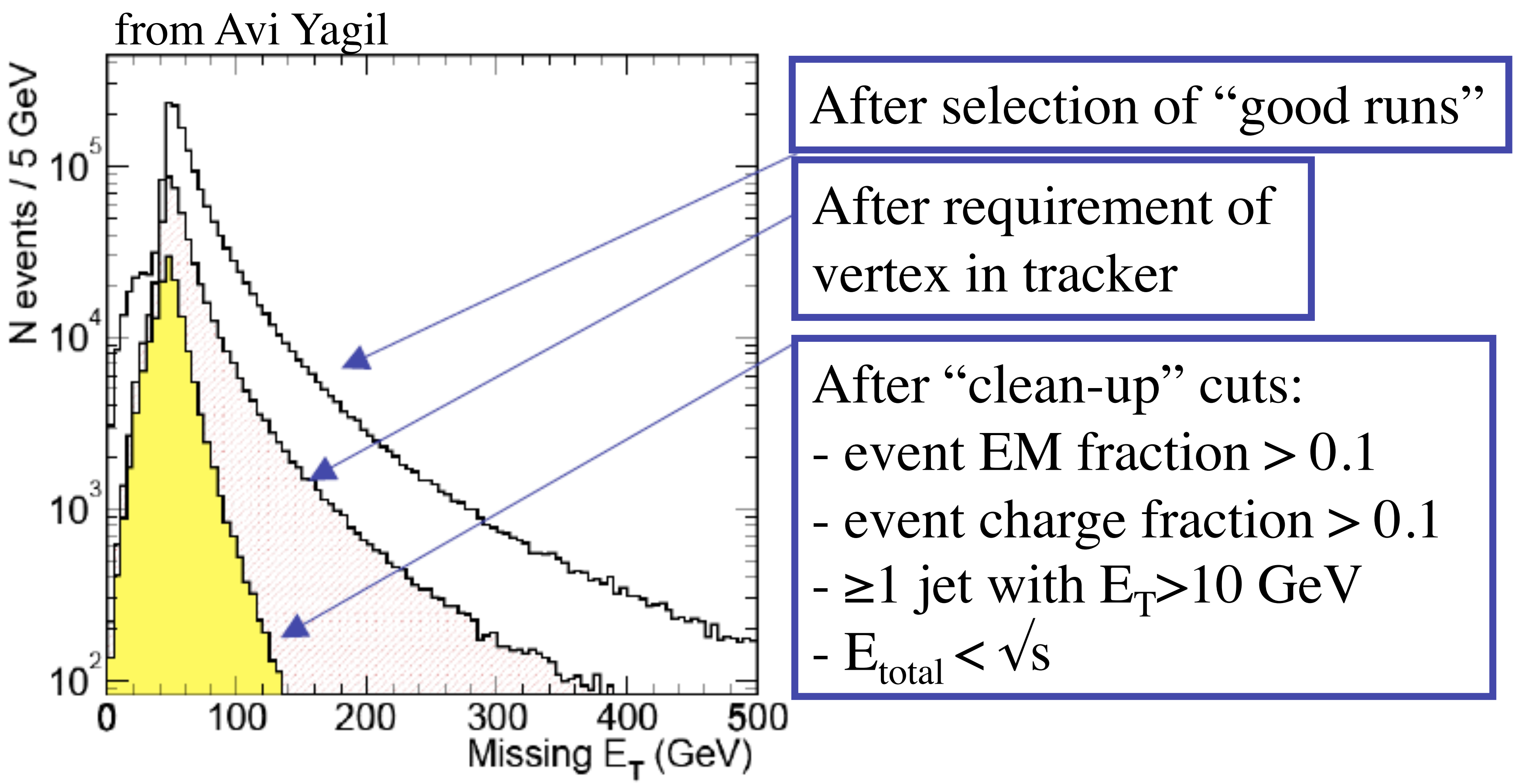
1 lepton



- "Effective Mass" = sum of p_T of all objects
- Similar and great (!) sensitivity in both modes
- Main backgrounds: top, W/Z+jets, QCD multi-jet

But how do we know the backgrounds!?!

Instrumental Backgrounds

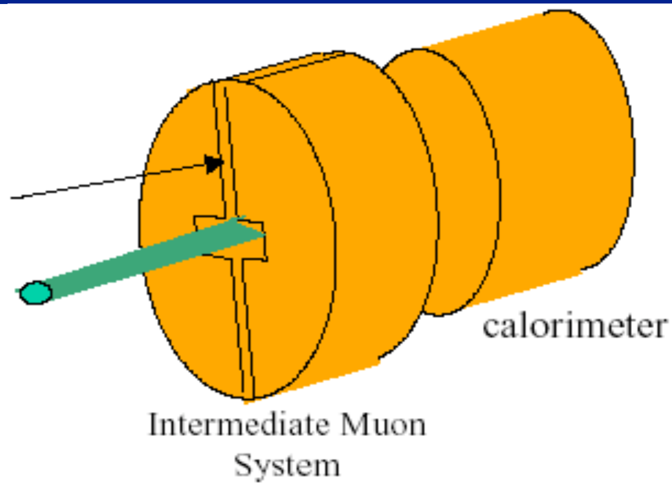


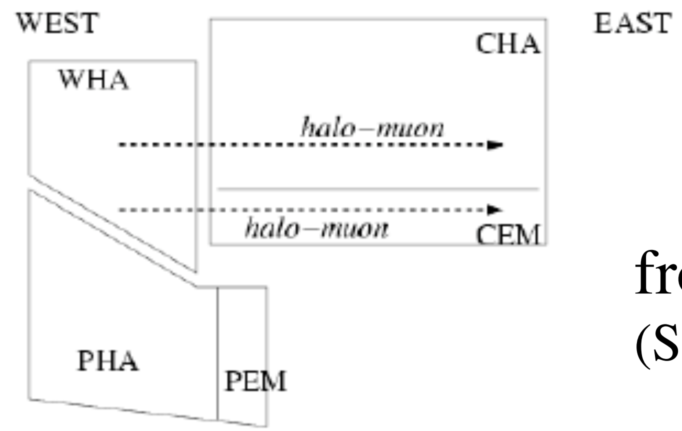
- Missing E_T distribution subject to many experimental effects
  - “If anything goes wrong it will affect missing E_T”

Sources of Instrumental Background

- Calorimeter Noise
  - Hot cells / coherent noise
    - Usually localized and can be rejected
- Calorimeter dead regions
  - Should only happen rarely in some runs
    - Should be removed by DQ criteria
- Cosmic rays and beam halo muons showering hard in calorimeter
  - Usually have no vertex but can overlap with MinBias event
    - Then have small tracking activity compared to calorimeter activity
  - Shower often only in hadronic calorimeter
- Example handles:
  - Track/calorimeter matches
  - Is direction of missing energy uniform?

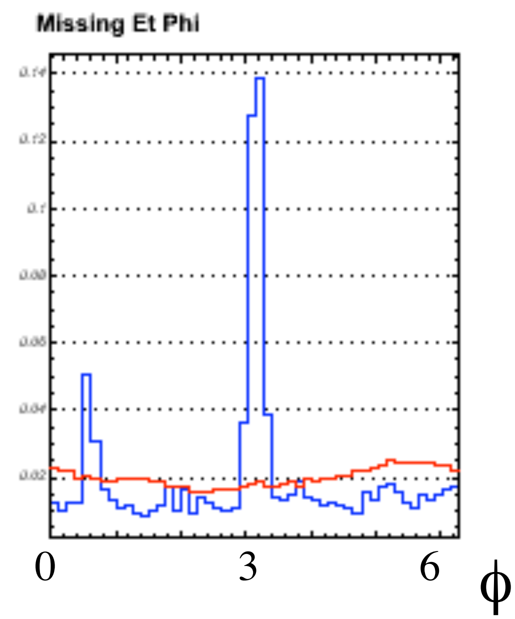
Beam-Halo Muon Background



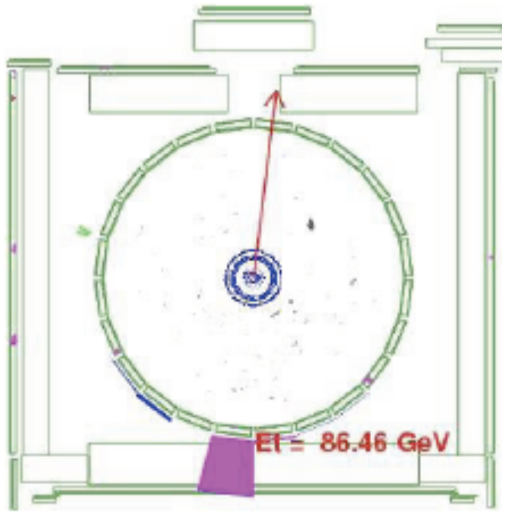


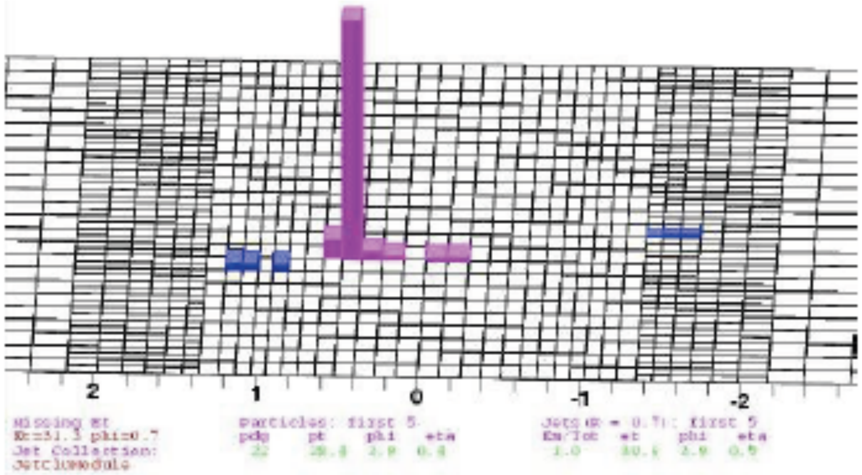
from CDF
(S.M.Wang)

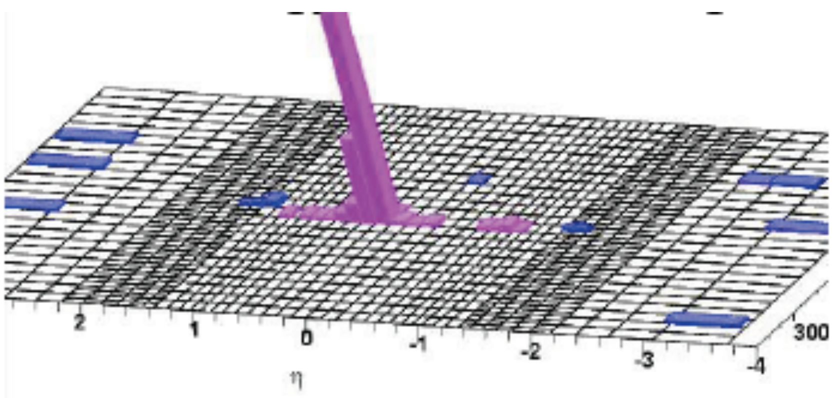
- Muon that comes from beam and goes through shielding
- Can cause showers in calorimeters
  - Shower usually looks not very much like physics jet
    - Often spike at certain azimuthal angles: π
  - But there is lots of those muons!
  - Can even cause problem for trigger rate



Some Cosmics and Beam-halo events





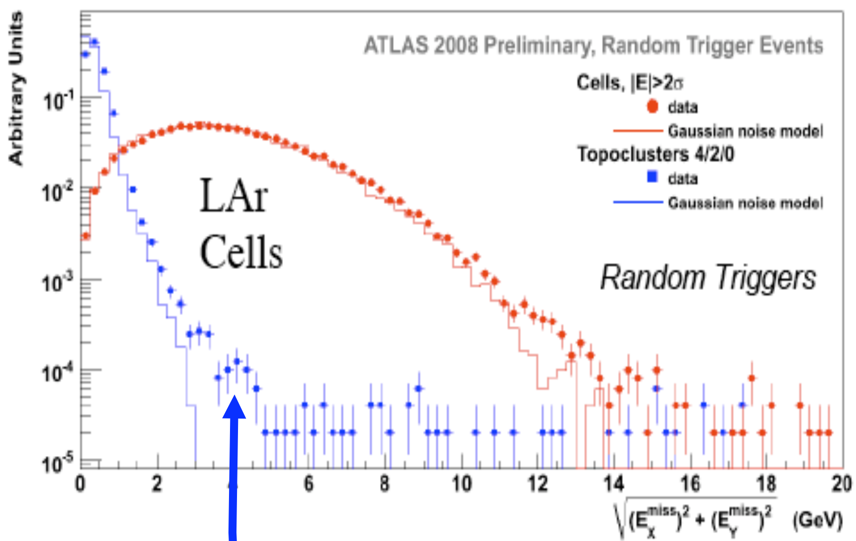


- Bigger problem for mono-jet than for multi-jet searches
- Can use
  - topological filters to reject events
  - Track matching calorimeter cluster

Instrumental Background: Studies with Cosmics

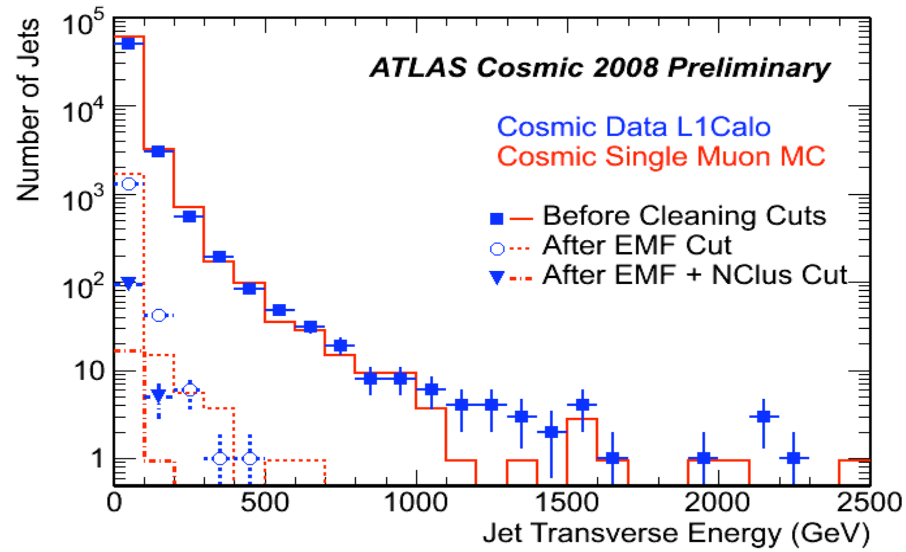
- Can learn a lot from cosmic ray data taking
  - ATLAS and CMS took cosmics for several weeks of running

2008 data: noise in random trigger



Noise in presampler:
Fixed in 2009

Developing cuts against cosmic Ray background



Amazing how well these properties are modeled by the cosmics simulation

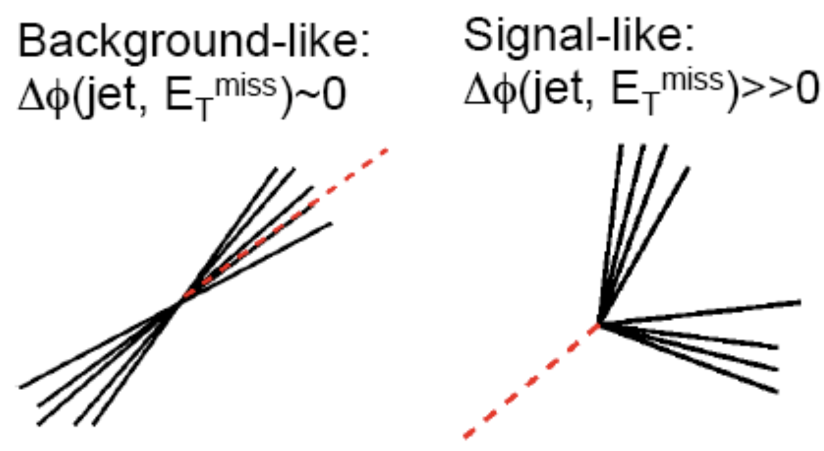
Physics Backgrounds

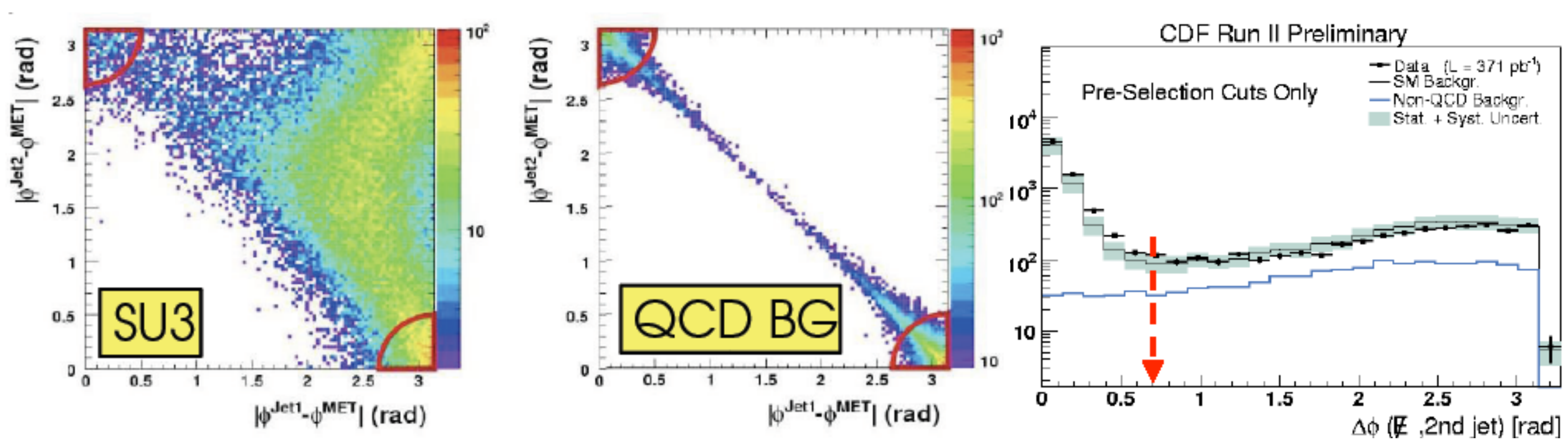
- QCD multi-jet (mosty for 0-lepton case)
  - Missing E_T due to
    - Poor jet resolution / cracks in calorimeters
    - Neutrino momentum in semi-leptonic b/c- decays
- W/Z+jets
  - Missing E_T due to ν's from $Z \rightarrow \nu\nu$, $W \rightarrow l\nu$
- Top
  - Missing E_T due to ν's from $tt \rightarrow WbWb \rightarrow l\nu + X$

How do we estimate them?

QCD Multi-jet

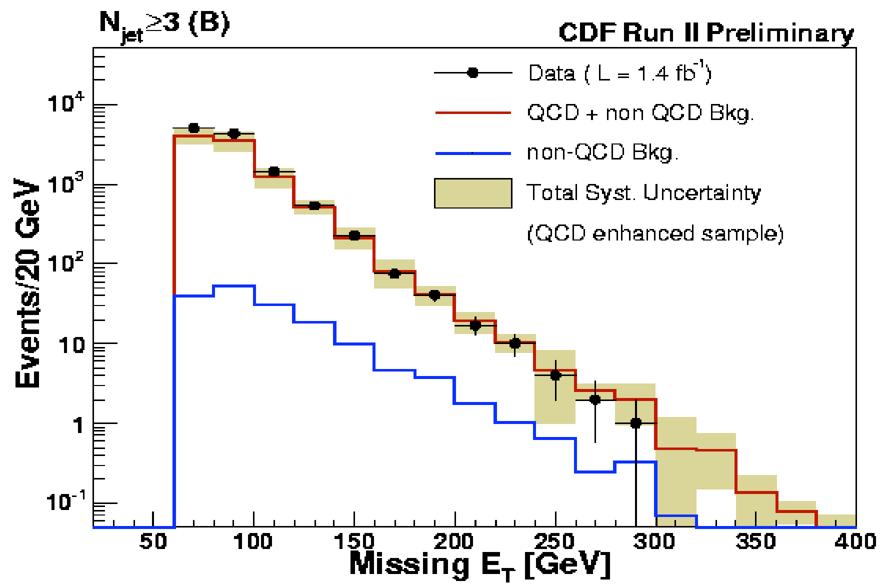
- Require large $\Delta\phi$
  - Between missing E_T and jets and between jets
  - Suppresses QCD dijet background due to jet mismeasurements



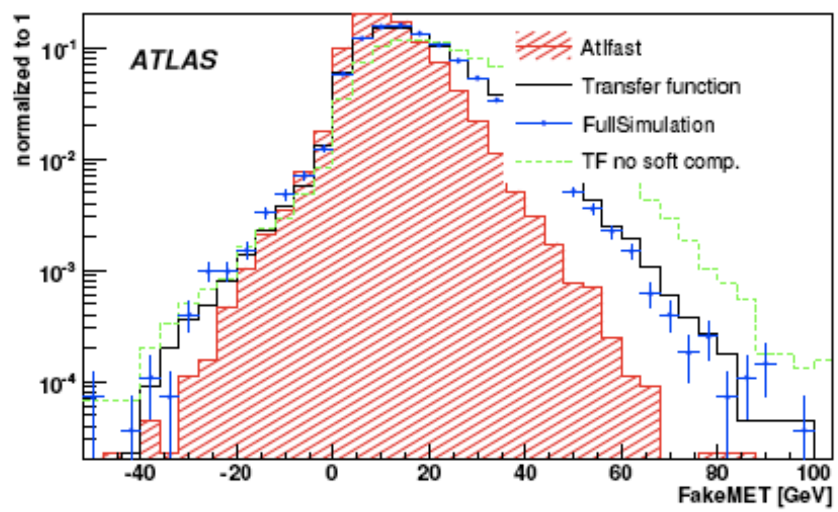


Methods to estimate remaining QCD multi-jet Background

1. CDF uses MC
   - Validate in region of low $\Delta\Phi$and low MET
   - Extrapolate to large using MC
   - Problem:
     - Relies on full MC simulation which can take “forever”



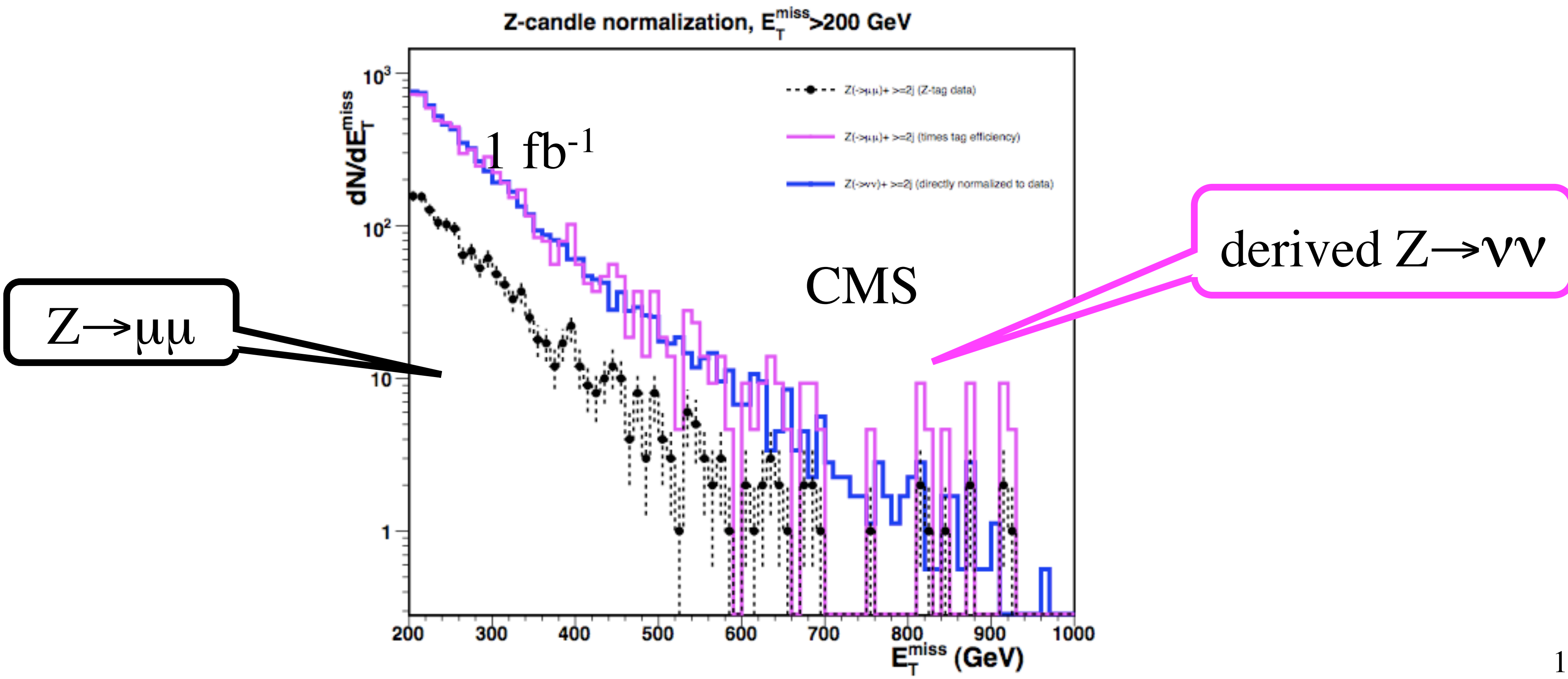
2. Parameterize truth jets with response function from full simulation
   - Validate against full simulation
   - Validate in region of lower MET
   - Advantage:
     - Do not need to simulate as many events
   - Need to make sure though that parameterization is really working



Using Z(→ll)+jets for estimating W/Z+jet background

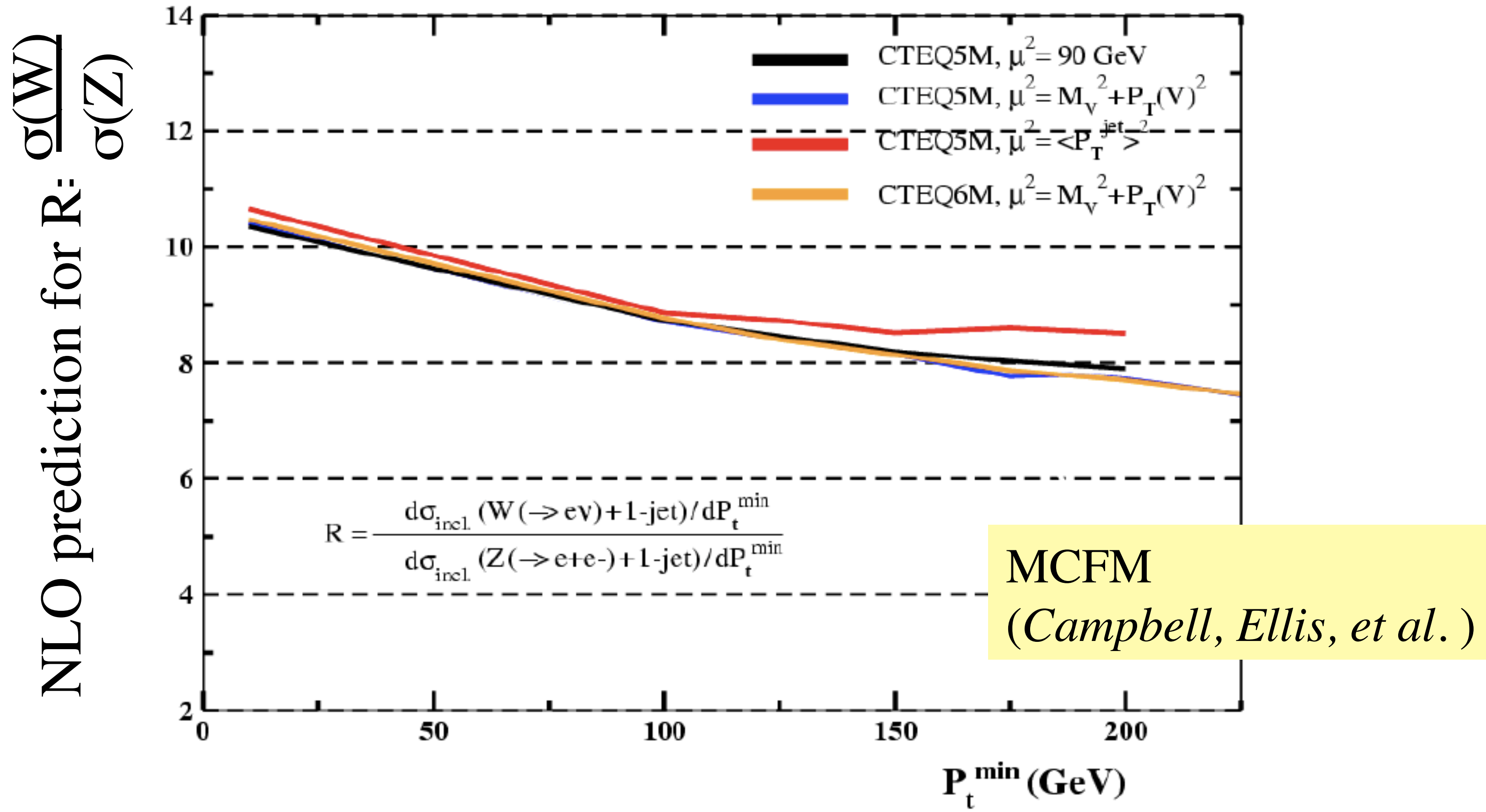
- Use Z(→ll)+jets to extrapolate to Z(→νν)+jets
  - $ME_T \sim p_T(Z)$

$$N_{Z\to\nu\bar{\nu}}(E_T^{\text{miss}}) = N_{Z\to\ell^+\ell^-}(p_T(\ell^+\ell^-)) \times c_{\text{Kin}}(p_T(Z)) \times c_{\text{Fidu}}(p_T(Z)) \times \frac{\text{Br}(Z\to\nu\bar{\nu})}{\text{Br}(Z\to\ell^+\ell^-)},$$



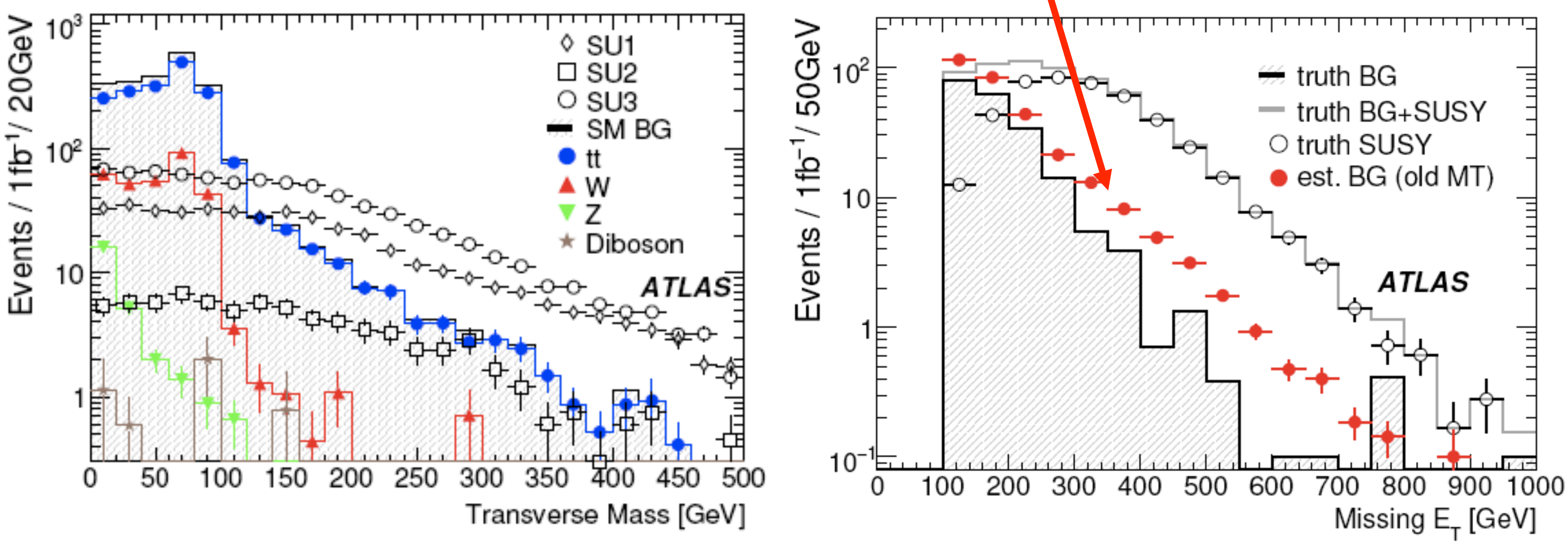
W+jets background estimate

- Use Z->ll +jets also for this background too
  - Rely on theoretical prediction for W+jets vs Z+jets
    - This is well known though (<15%)!



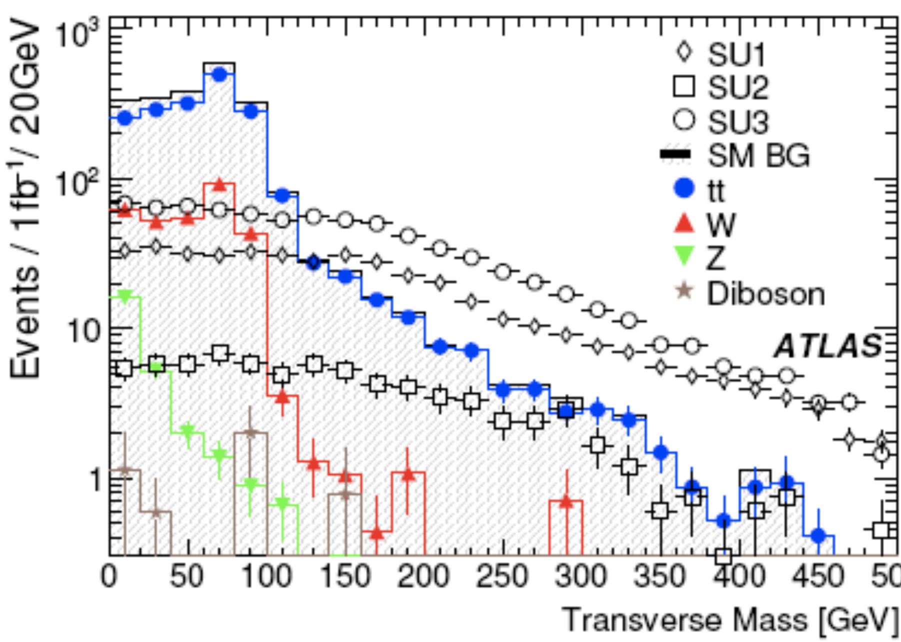
Top and W+jets background estimate

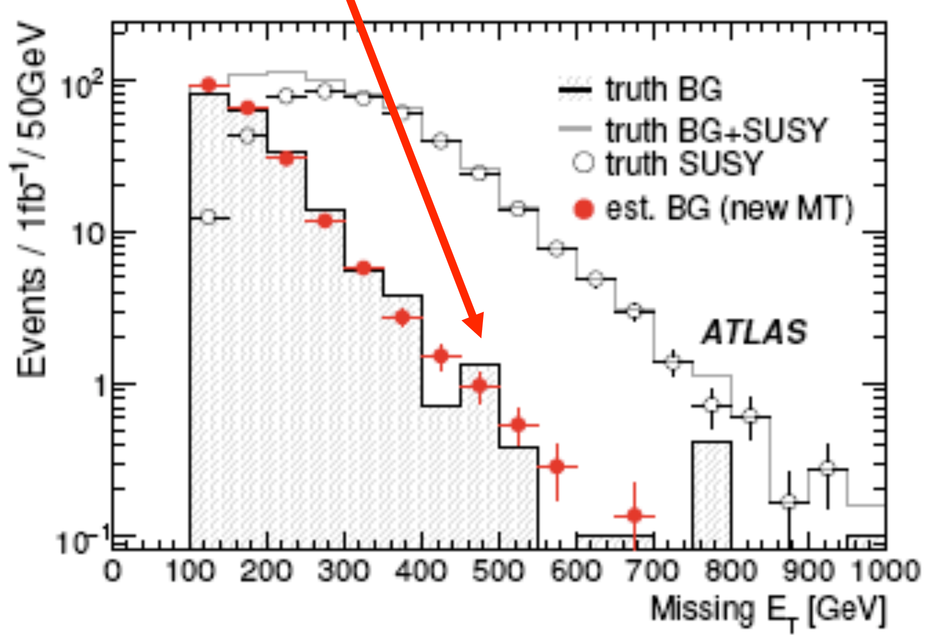
- Use region of low $m_T(W)$
  - Extrapolate to signal region using MC
  - But may be contaminated by SUSY => overestimate BG
    - depending on specifics of model



Top and W+jets background estimate

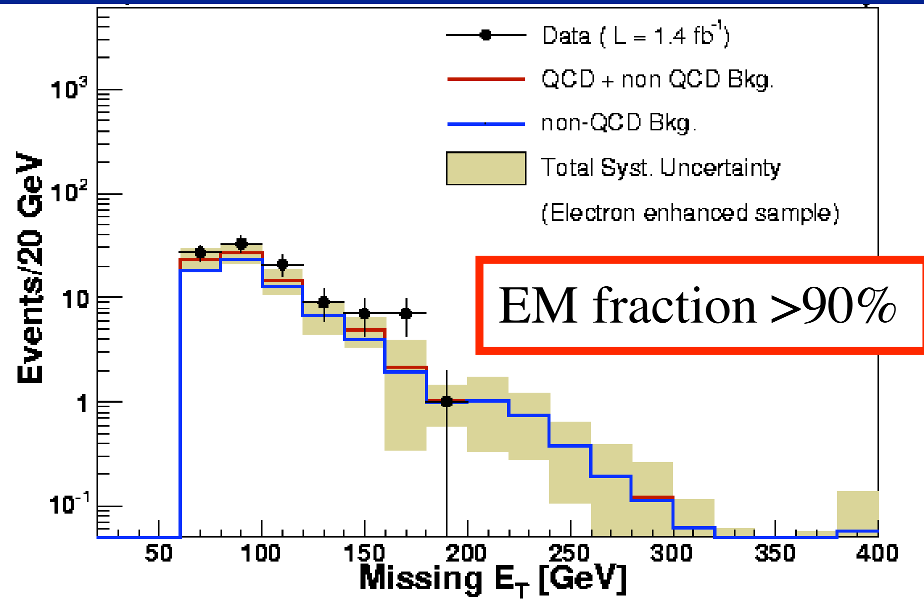
- Use region of low $m_T(W)$
  - Extrapolate to signal region using MC
  - But may be contaminated by SUSY => overestimate
    - depending on specifics of model
  - Can attempt “SUSY background subtraction” to correct for it

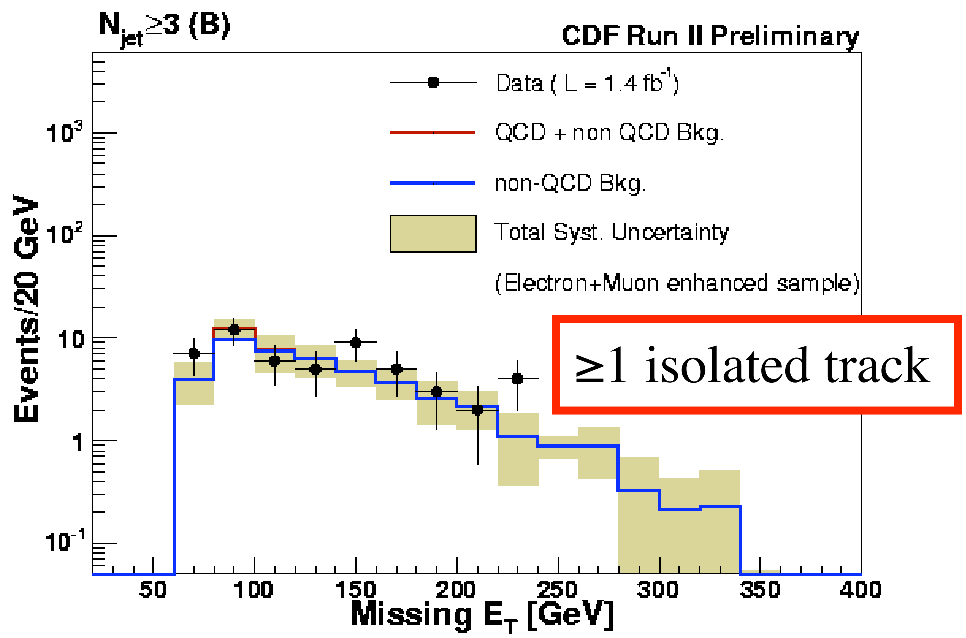




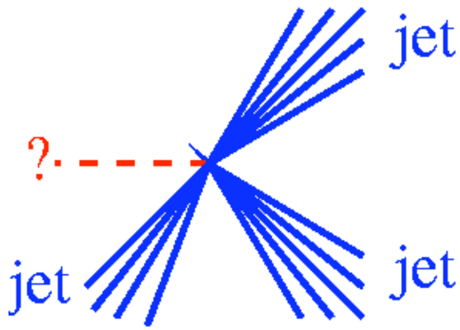
W+jets, Z+jets and Top background

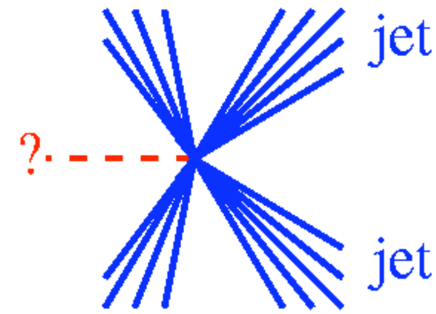
- Checks at Tevatron 0-lepton analysis
  - Background sources:
    - W/Z+jets, top
    - Suppressed by vetoes:
      - Events with jet with EM fraction>90%
        - Rejects electrons
      - Events with isolated track
        - Rejects muons, taus and electrons
  - Define control regions:
    - W/Z+jets, top
      - Make all selection cuts but invert lepton vetoes
    - Gives confidence in those background estimates
      - Modeled using Alpgen MC
      - Cross sections determined using NLO calculation
- May not work at LHC due to expectation of large cascade decays

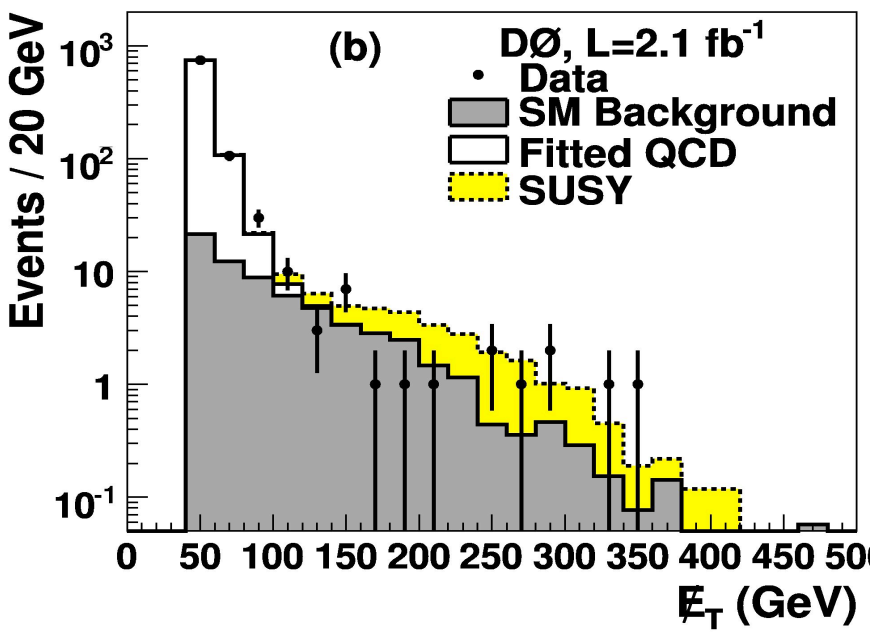


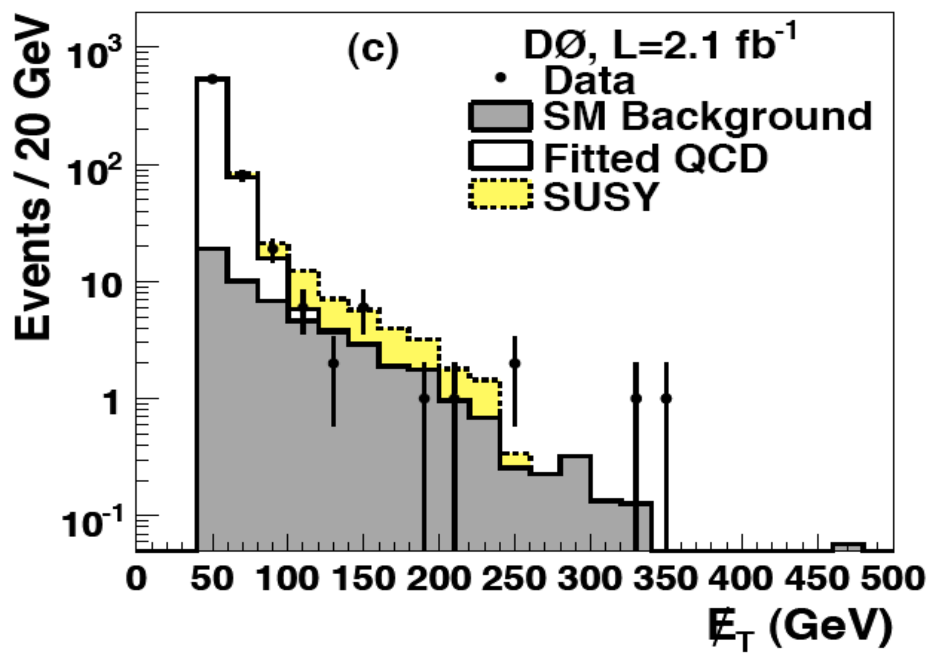


Final Analysis Plots at the Tevatron



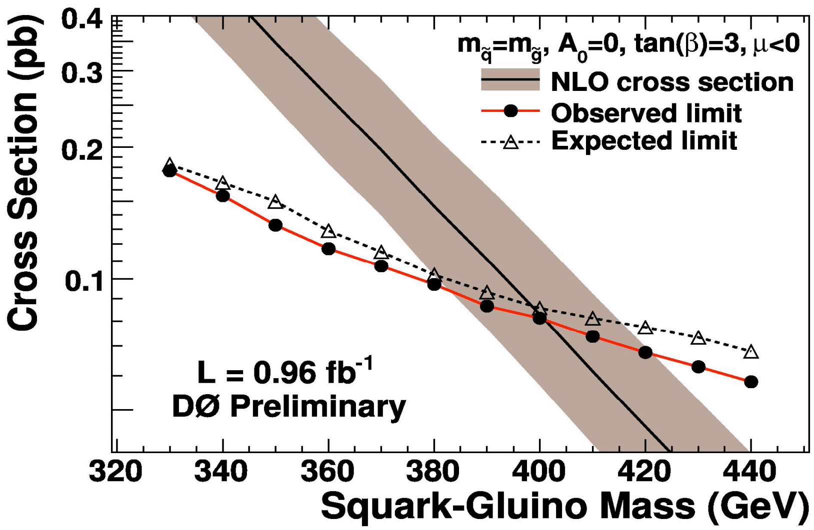


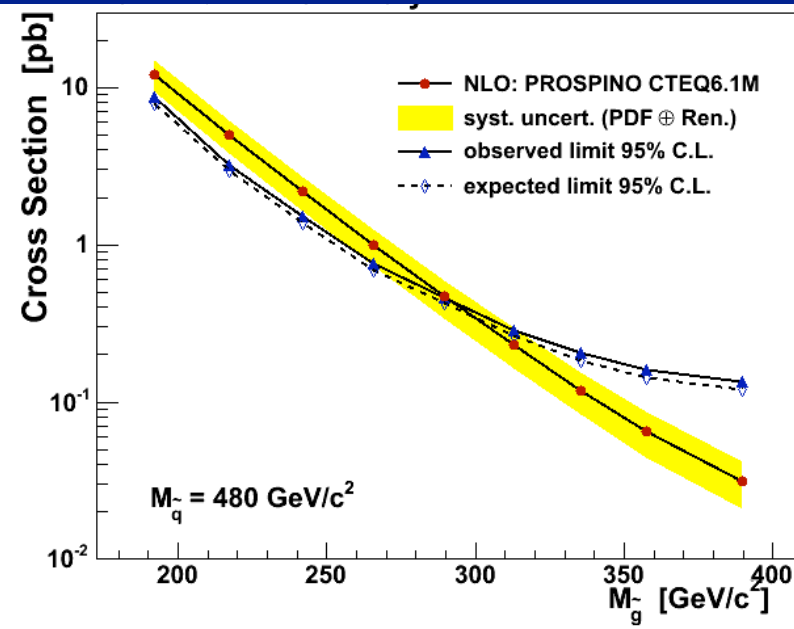




Data agree with background estimate => derive limits

Cross Section Limits

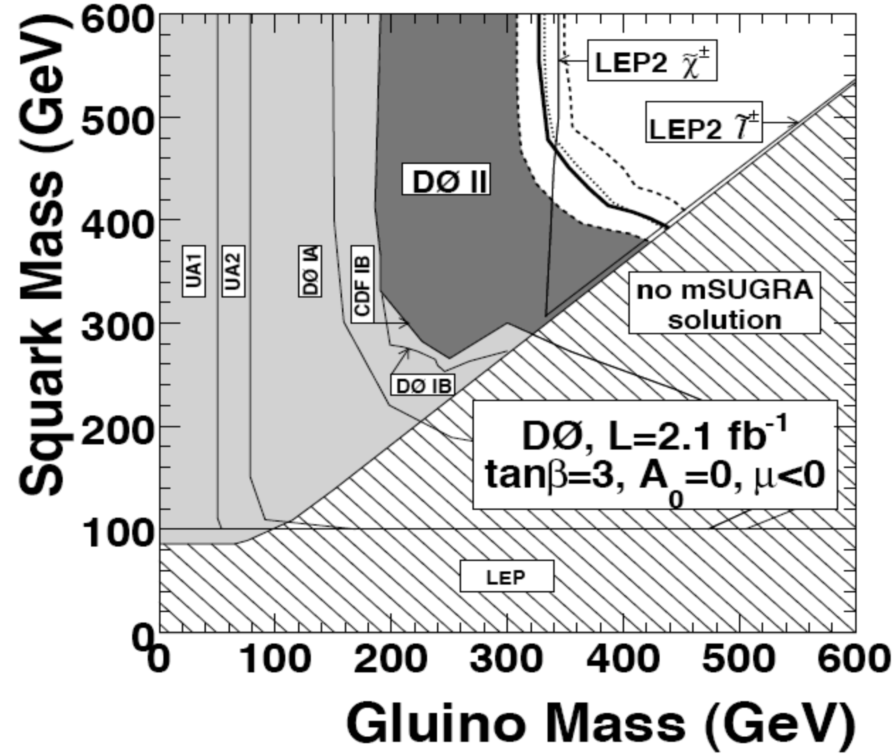


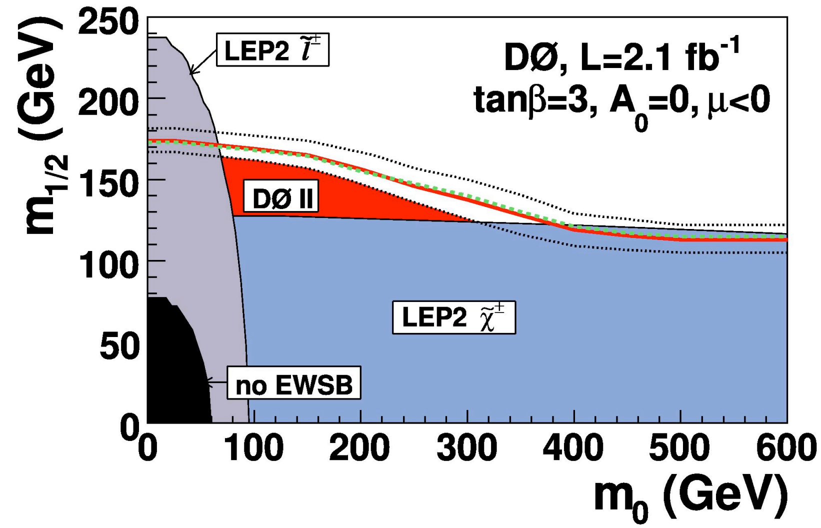


- No excess in data
  - Evaluate upper limit on cross section
  - Find out where it crosses with theory
- Theory has large uncertainty: ~30%
  - Crossing point with theory lower bound ~ represents limit on squark/gluino mass

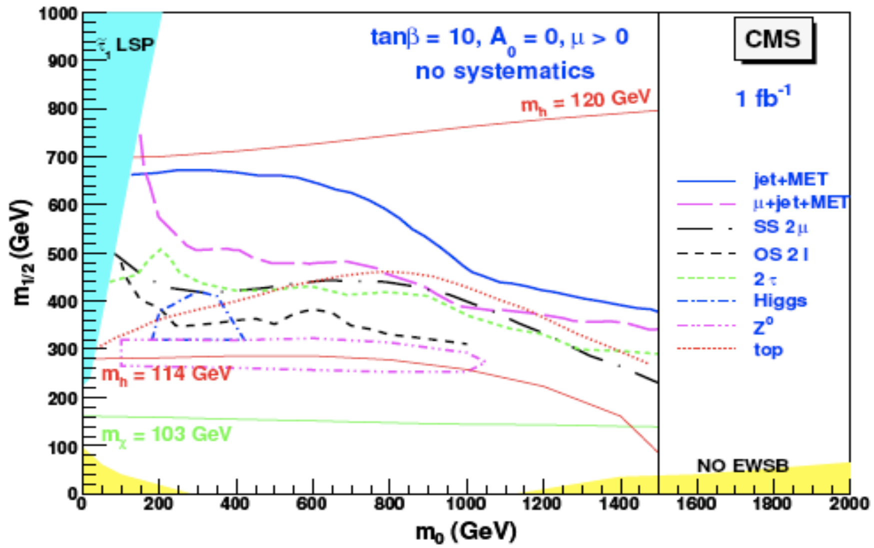
Squark and Gluino Mass Limits

- Set constraints on masses at EWK scale:
  - $M(\tilde{g})>308$ GeV
  - $M(\tilde{q})>379$ GeV
- Can also represented in terms of GUT scale parameters
  - Within constrained models

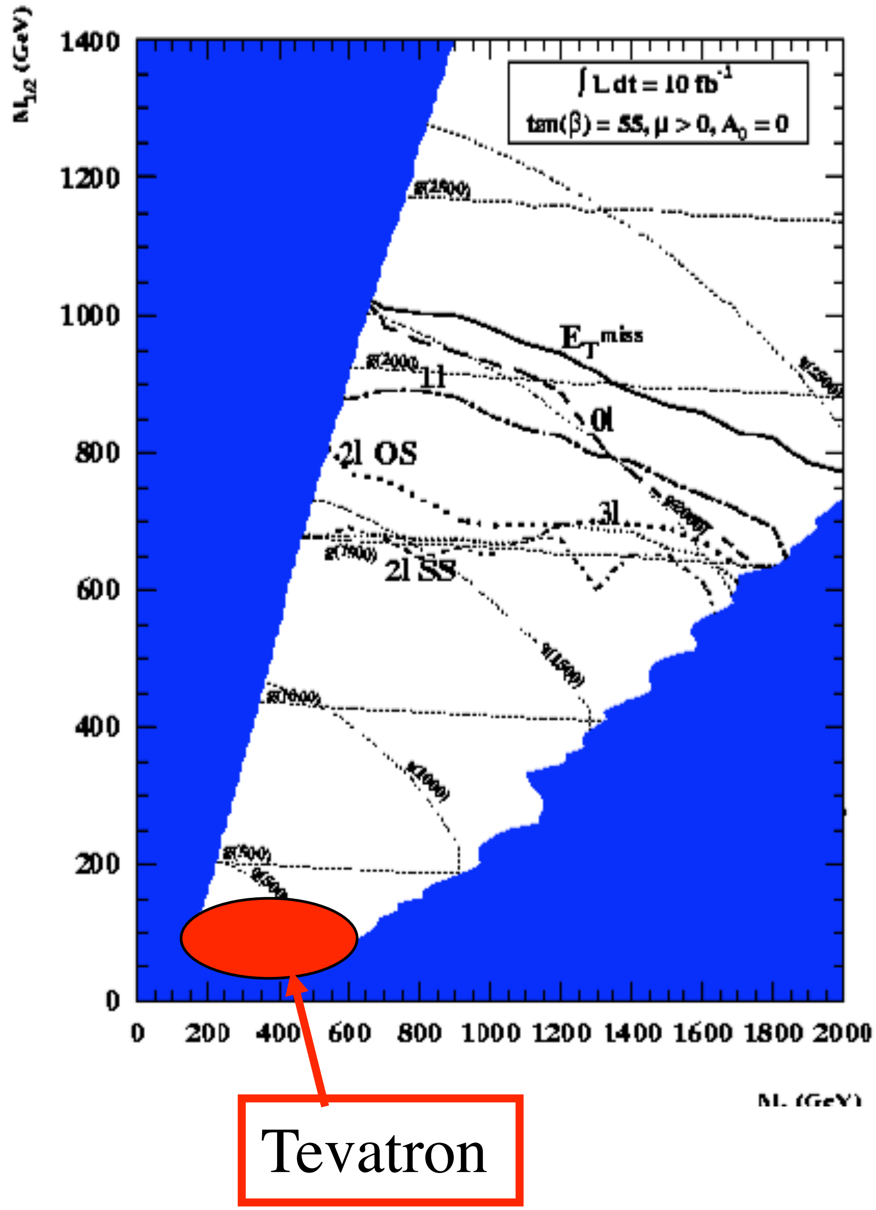


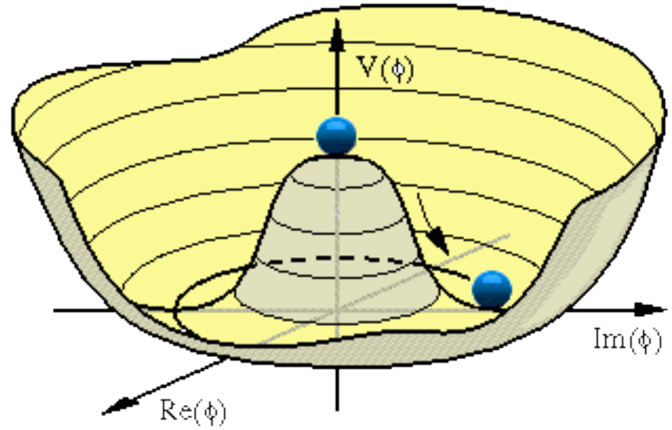


LHC SUSY Discovery Reach

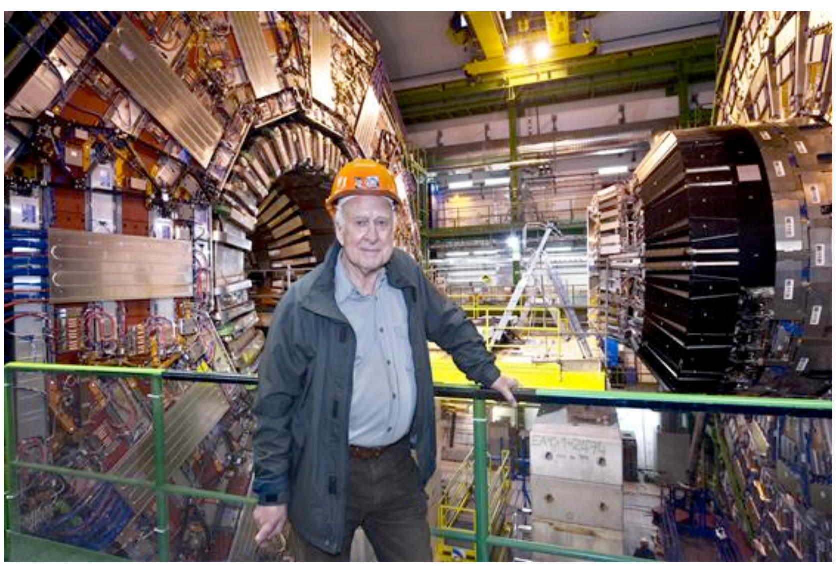


- With 1 fb^{-1}:
  - Sensitive to $m(\tilde{g})$<1000 GeV/c^2
- With 10 fb^{-1}:
  - Sensitive to $m(\tilde{g})$<1800 GeV/c^2
- Amazing potential!
  - If data can be understood
  - If current MC predictions are ≈ok



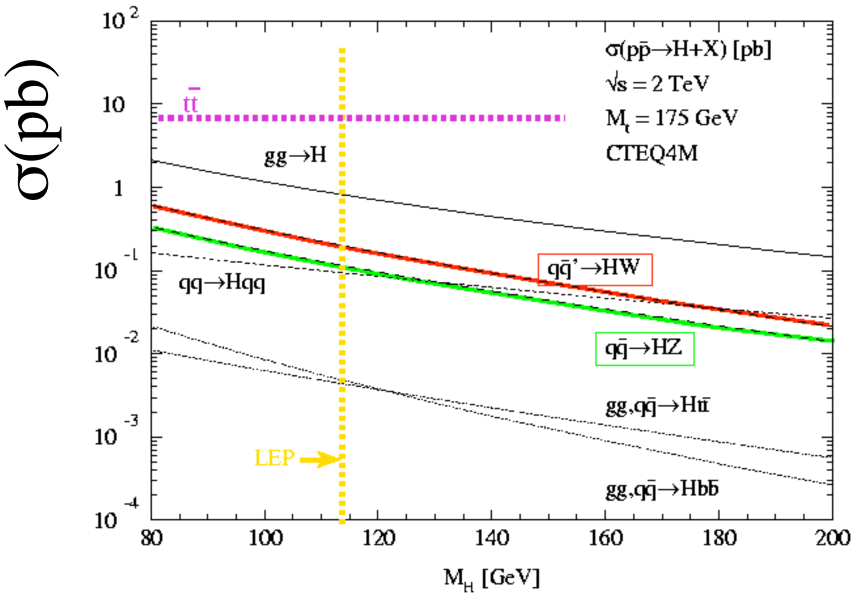


The Higgs Boson

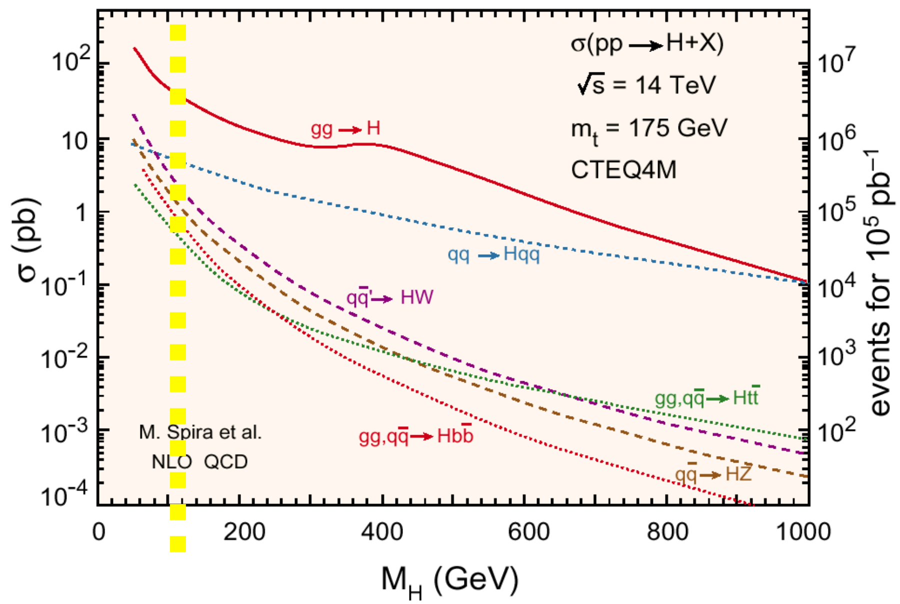


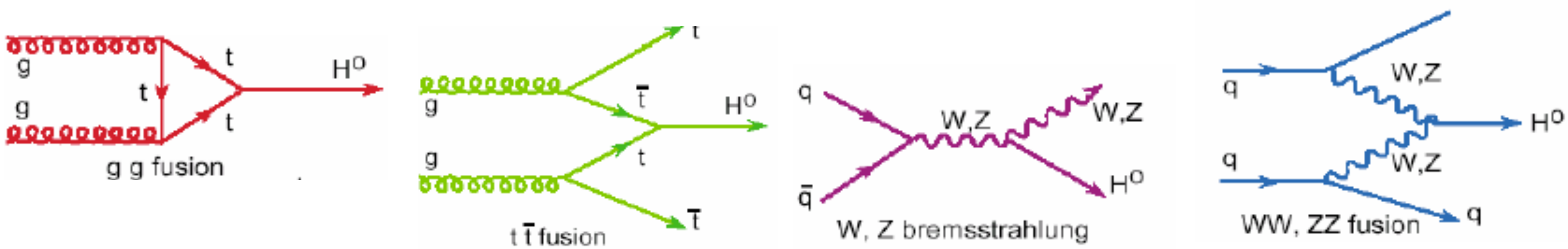
Higgs Production: Tevatron and LHC

Tevatron



LHC

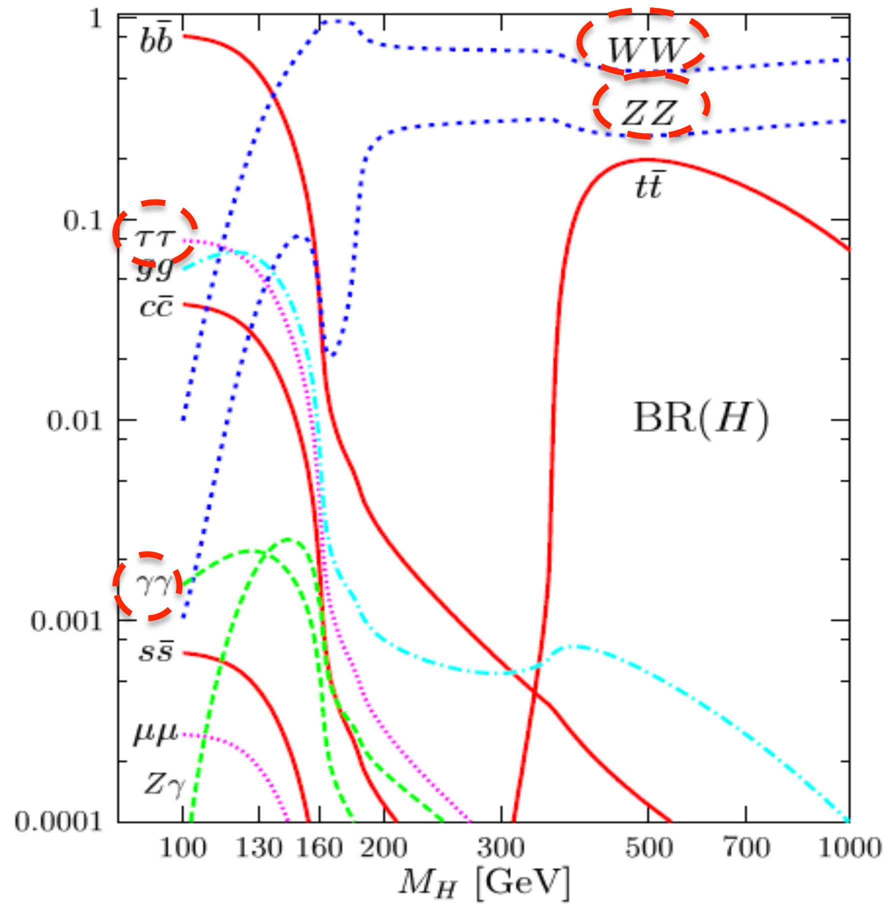




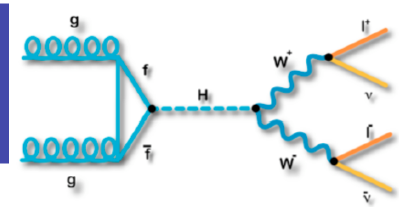
dominant: gg→ H, subdominant: HW, HZ, Hqq

Higgs Boson Decay

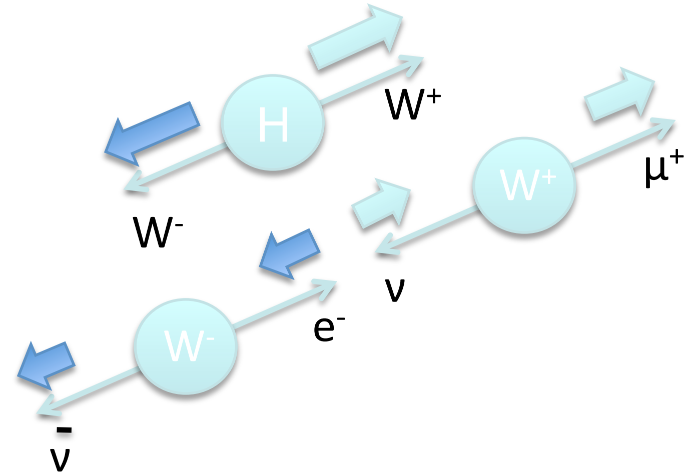
- Depends on Mass
- M_H<130 GeV/c^2:
  - $b\bar{b}$ dominant
  - WW and $\tau\tau$ subdominant
  - $\gamma\gamma$ small but useful
- M_H>130 GeV/c^2:
  - WW dominant
  - ZZ cleanest

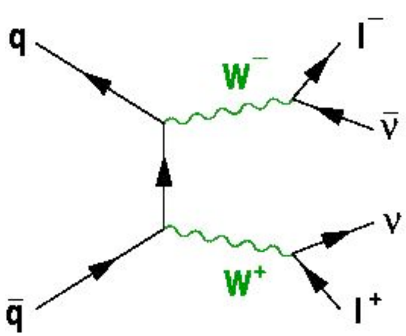


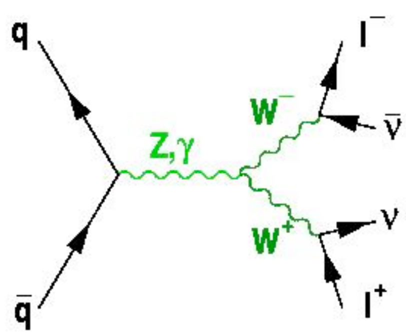
$H \rightarrow WW^{(*)} \rightarrow l^+l^-\bar{\nu}\nu$

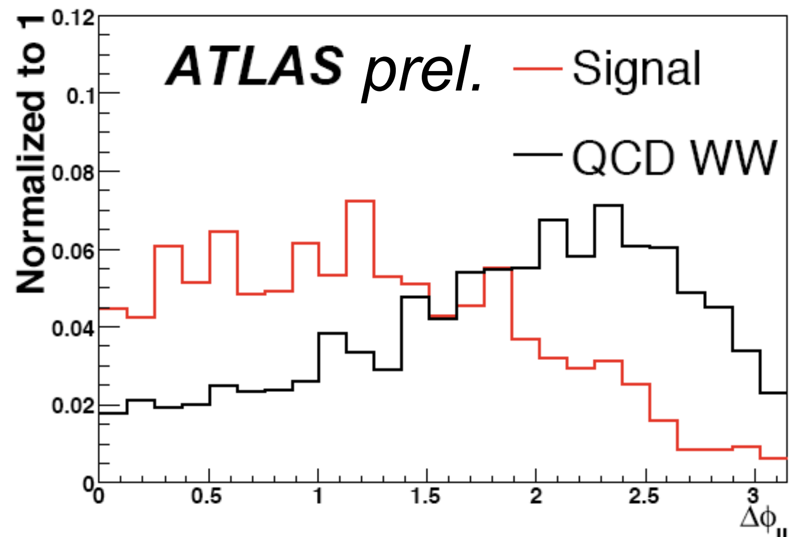


- Higgs mass reconstruction impossible due to two neutrinos in final state
- Make use of spin correlations to suppress WW background:
  - Higgs is scalar: spin=0
  - leptons in $H \rightarrow WW^{(*)} \rightarrow l^+l^-\nu\nu$ are collinear
- Main background:
  - WW production





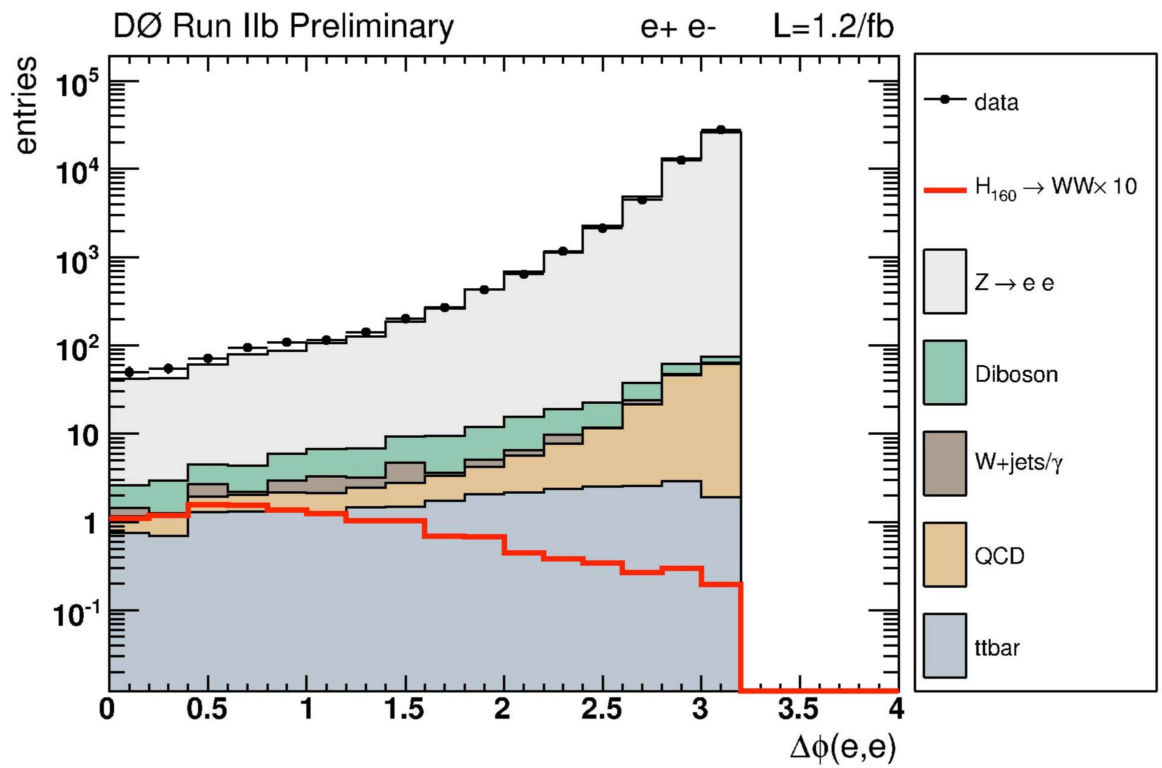




$H \to WW^{(*)} \to l^+l^-\nu\nu$ ($l=e,\mu$)

- **Event selection:**
  - 2 isolated e/μ :
    - p_T > 15, 10 GeV
  - Missing E_T >20 GeV
  - Veto on
    - Z resonance
    - Energetic jets

- **Main backgrounds**
  - SM WW production
  - Top
  - Drell-Yan
  - Fake leptons



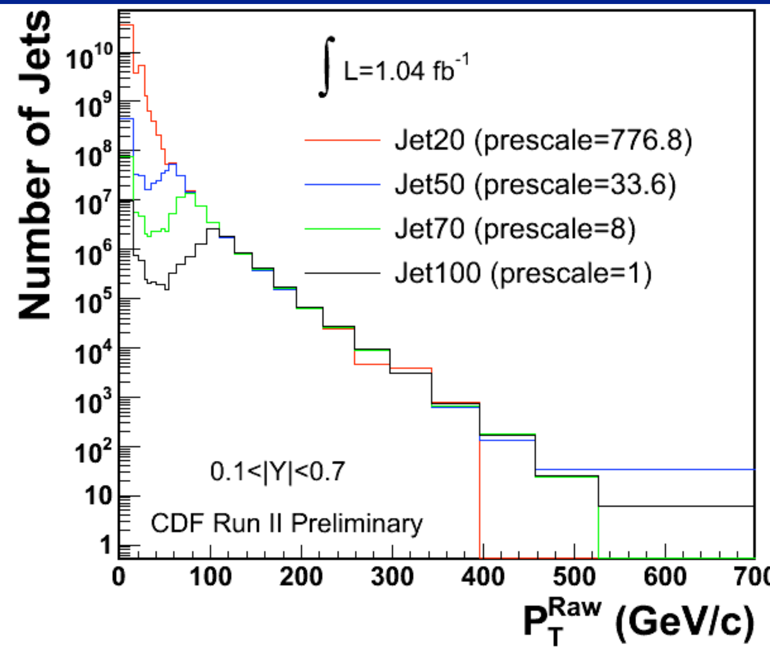
- **Plot everything under the sun**
  - to convince yourself you have the background right

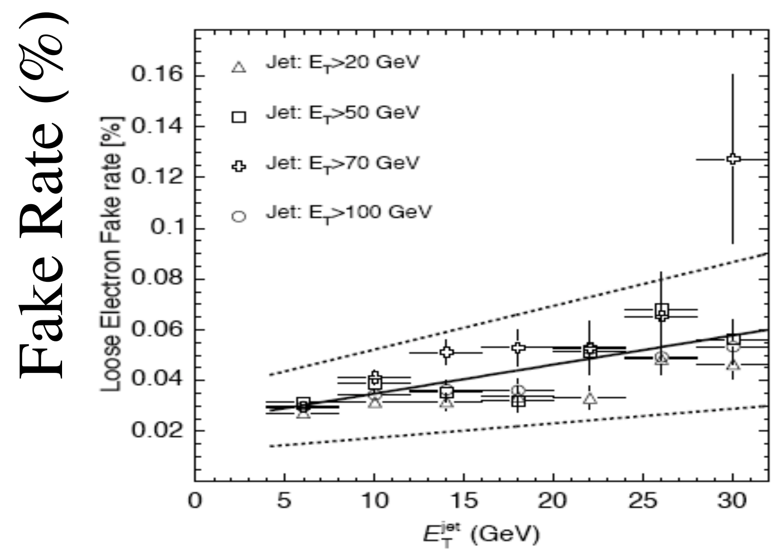
Jets faking Electrons

- Jets can pass electron ID cuts,
  - Mostly due to
    - early showering charged pions
    - Conversions: $\pi^0 \to \gamma\gamma \to ee+X$
    - Semileptonic b-decays
  - Difficult to model in MC
    - Hard fragmentation
    - Detailed simulation of calorimeter and tracking volume
- Measured in inclusive jet data at various E_T thresholds
  - Prompt electron content negligible:
    - N_{jet}~10 billion at 50 GeV!
  - **Fake rate per jet:**

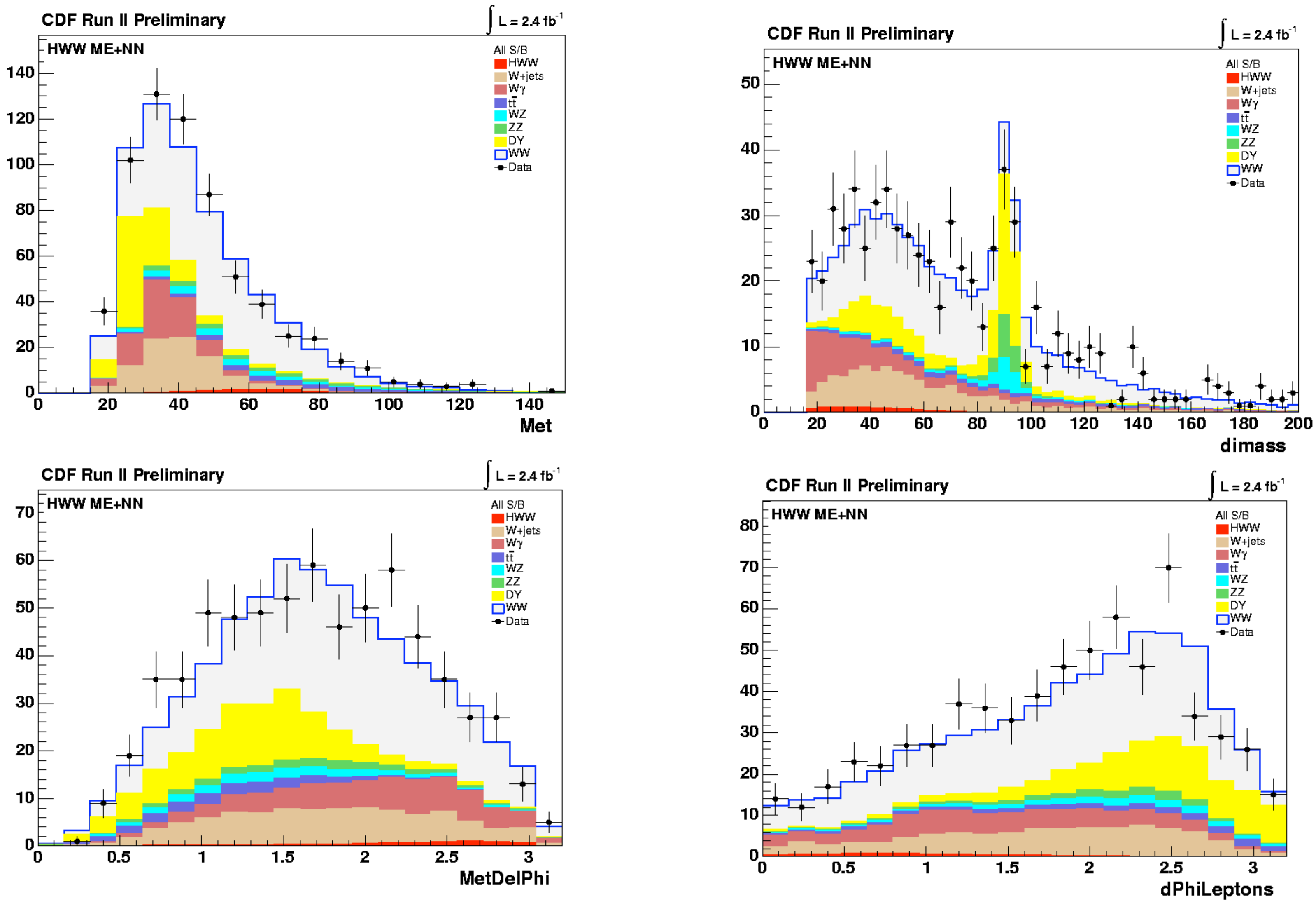
	CDF	ATLAS
Loose cuts	5×10^{-4}	5×10^{-3}
Tight cuts	1×10^{-4}	1×10^{-5}

  - Typical uncertainties 50%





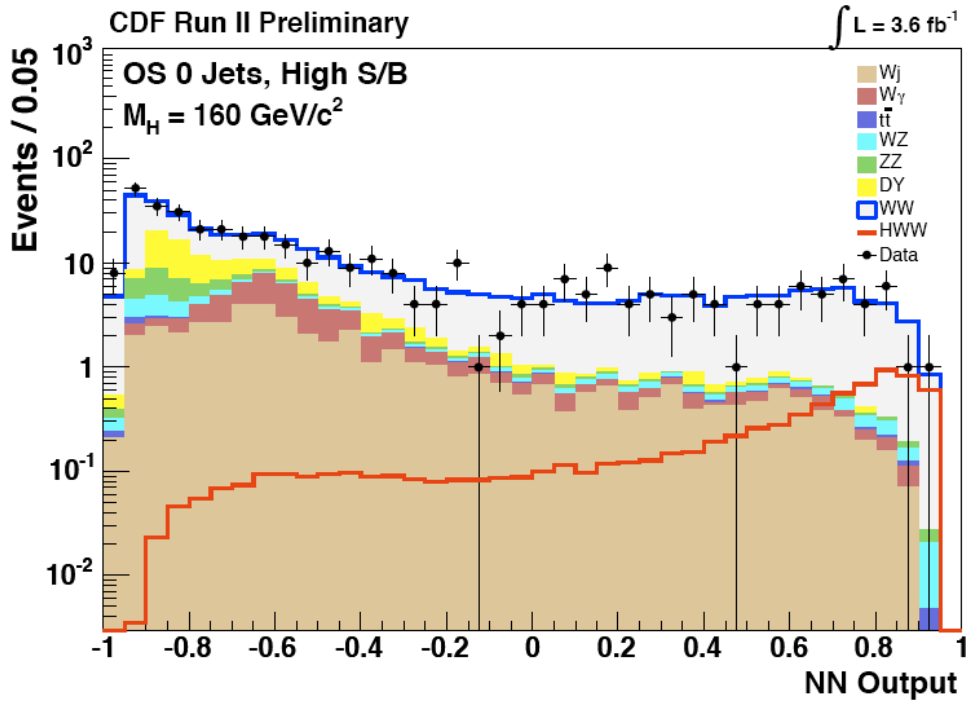
Plot Everything Under the Sun..



- **Validates the background prediction**
  - **Very often these plots "don't work" since there is some problem**
  - **Now plug all into sophisticated techniques!**

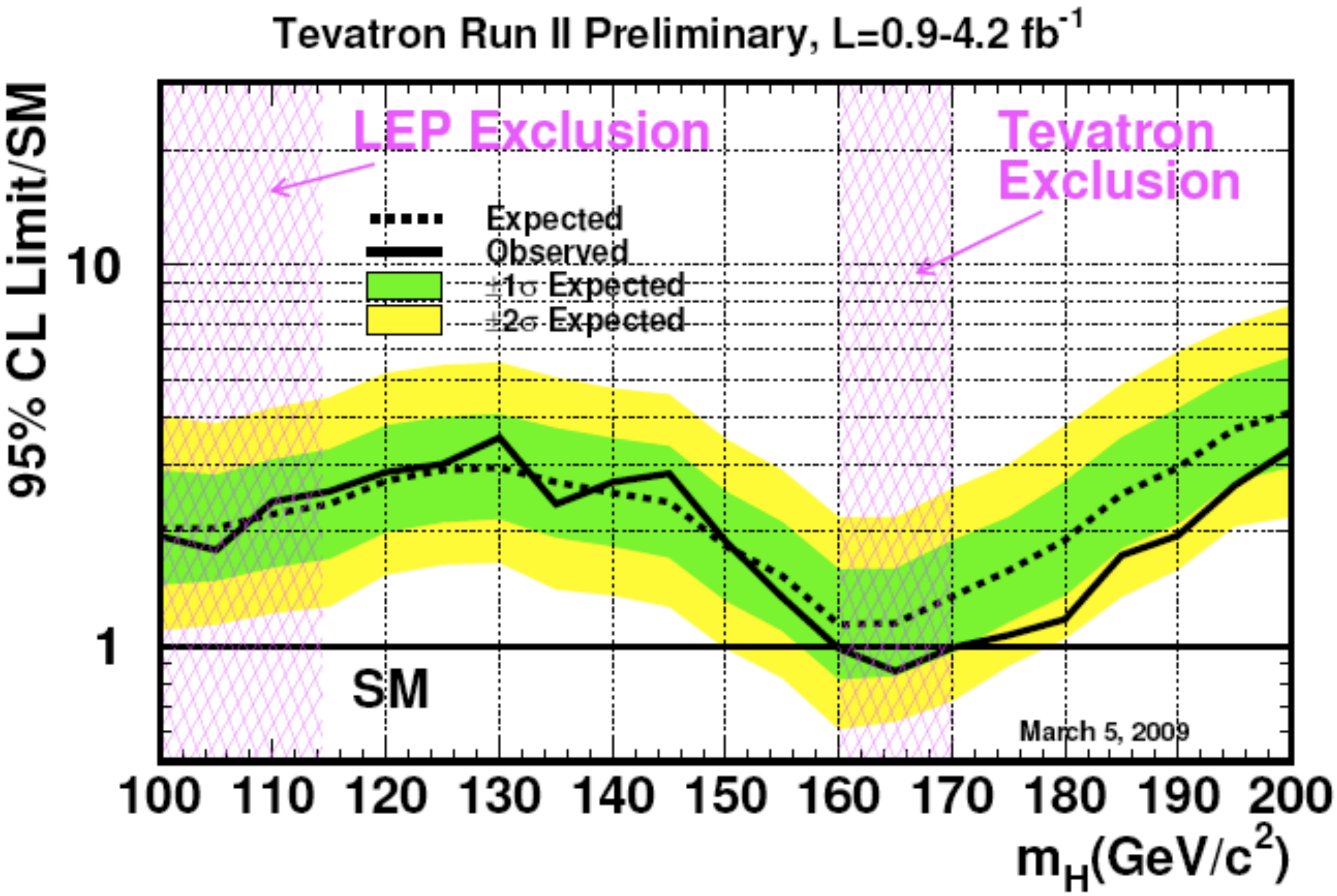
NN Output

$M_H = 160$ GeV/c^2	
$t\bar{t}$	1.35 ± 0.21
DY	80 ± 18
WW	318 ± 35
WZ	14 ± 1.9
ZZ	20.7 ± 2.8
W+jets	113 ± 27
$W\gamma$	92 ± 25
Total Background	**637 ± 67**
$gg \to H$	9.5 ± 1.4
Total Signal	**9.5 ± 1.4**
Data	**654**



- Data agree well with background hypothesis
  - S/B ~0.3 at high NN values

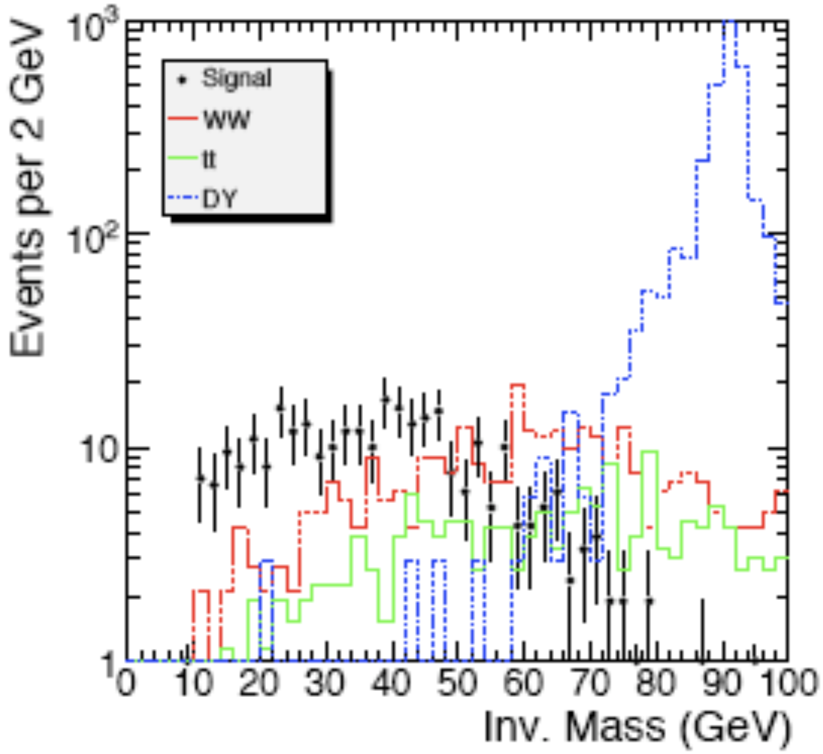
Higgs Cross Section Limit

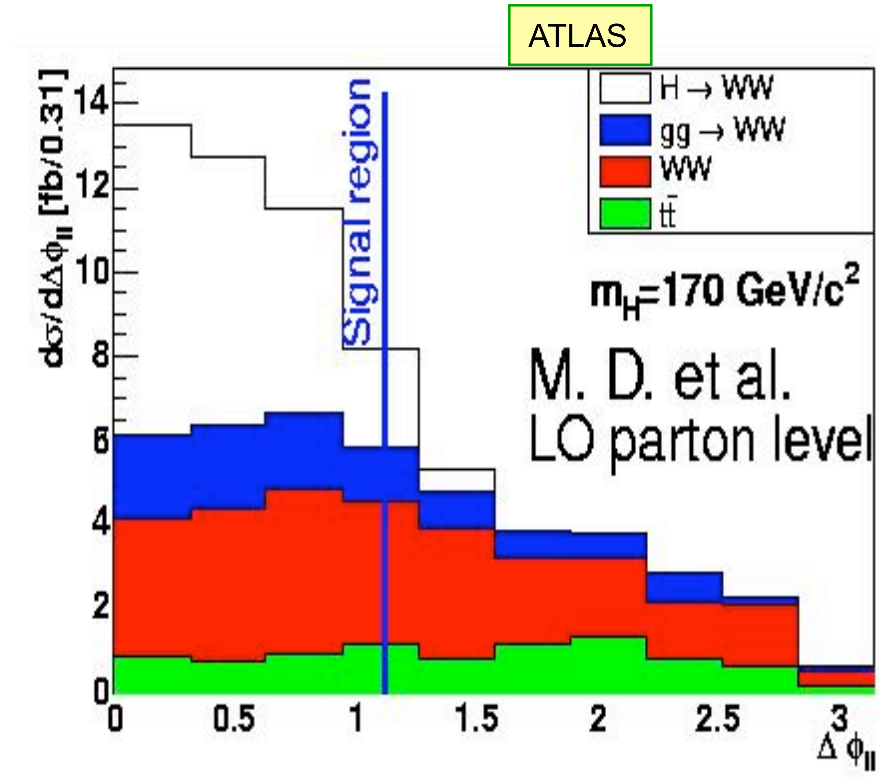


- $160 < m_H < 170$ GeV excluded at 95% C.L.
  - Note that the limit is ~1σ better than expected
- For m_H=120 GeV: $\sigma_{limit} / \sigma_{SM} = 2.8$

Early Higgs Signals at LHC

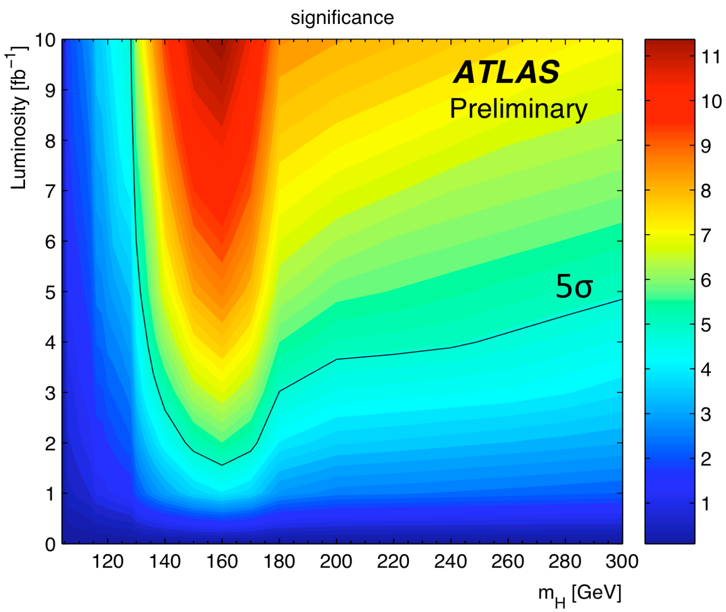
$H \rightarrow WW^*$ (m_H=170 GeV)

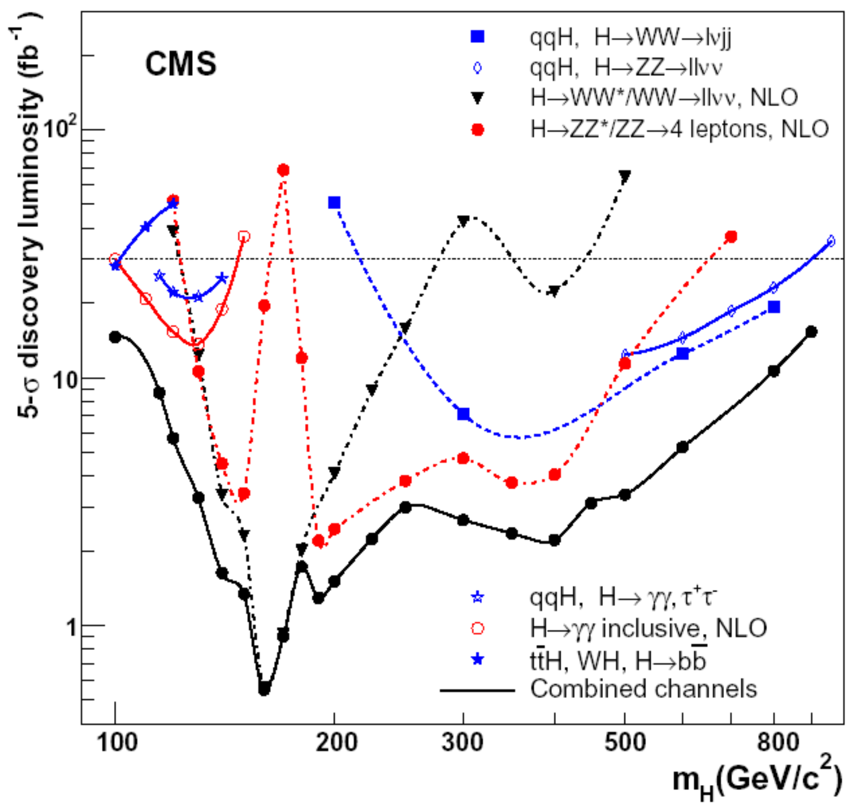


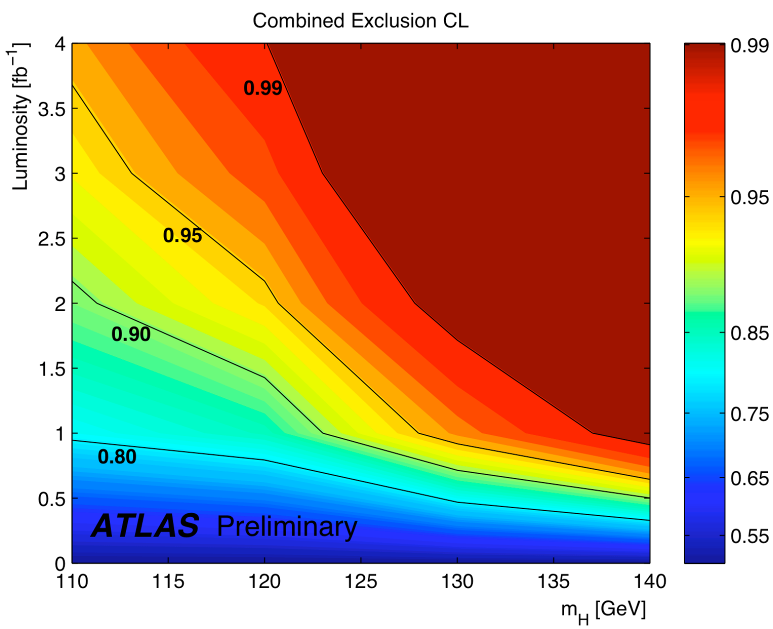


LHC has about 4 times better
signal / background than Tevatron

LHC SM Higgs Discovery Potential







- 5σ discovery over full mass range with ~20 fb^{-1}
  - Most challenging at low mass
- 95% exclusion over full mass range with ~4 fb^{-1}

Conclusions

- Background estimate most crucial aspect for searches
- LHC has an amazing discovery potential
  - Supersymmetry already with ~100 pb^{-1}
    - Also other high mass particles, e.g.
    - Z', Extra Dimensions, 4th generation quarks, …
  - Higgs boson: 1-10 fb^{-1}
- Let's hope that many exciting things will be found!!!

Some Remarks on Advanced Analysis Techniques

- **Quite a few techniques available:**
  - Neural Network, Likelihood, Boosted Decision Tree, Matrix Element, …
  - No clear winner has yet been identified
    - Some are more transparent than others
- **Why do we trust them less than simple analyses?**
  - Simple kinematic quantities can be calculated at NLO by theorists while e.g. NN distribution cannot
    - Gives confidence, good cross-check!
  - Techniques exploit correlations between variables
    - Harder to understand if the MC models correlations correctly
    - More validation needed (=> analysis takes longer)
  - Less transparent
    - Worry is always that it exploits some MC feature that does not reflect the data
- **Can and has been done of course though**
  - But only in mature experiments